PROCEEDINGS OF TDWG

Biodiversity Information Science and Standarts 1: e20637
doi: 10.3897/tdwgproceedings.1.20637

OPEN ACCESS

Conference Abstract

Using MIxS: An Implementation Report from Two Metagenomic Information Systems

Joel Sachs‡, Luke Thompson§, Nazir El-Kayssi‡, Satpal Bilkhu‡

‡ Agriculture and Agri-Food Canada, Ottawa, Canada
§ NOAA Southwest Fisheries Science Center, La Jolla, CA, United States of America

Corresponding author: Joel Sachs (joel.sachs@agr.gc.ca)

Received: 27 Aug 2017 | Published: 28 Aug 2017

Citation: Sachs J, Thompson L, El-Kayssi N, Bilkhu S (2017) Using MIxS: An Implementation Report from Two Metagenomic Information Systems. Proceedings of TDWG 1: e20637.
https://doi.org/10.3897/tdwgproceedings.1.20637

Abstract

MIxS (Minimum Information about any Sequence) (Yilmaz et al. 2011) is a metadata standard of the Genomics Standards Consortium (GSC), designed to make sequence data findable, accessible, and interoperable. It contains fields for recording physical and chemical characteristics of the sampling environment, geographical and habitat information, and other metadata about the sample and its provenance, which are critical for downstream intepretation of data derived from the sample. We will present our experience implementing MIxS in two metagenomic information systems – the Earth Microbiome Project (EMP) and the Government of Canada (GoC) Ecobiomics project.

The EMP (Gilbert et al. 2014) is an ongoing effort to crowdsource environmental microbiome samples from around Earth, then sequence and analyze them using a standardized workflow. The EMP has aggregated and sequenced over 50,000 samples, which are queryable using a publicly available catalogue. A meta-analysis of the first 25,000 samples is currently in review. MIxS and the Environment Ontology (ENVO) (Buttigieg et al. 2016) have been useful in structuring environmental metadata from EMP studies. For the particular application of the EMP meta-analysis, however, several issues were encountered: often there are multiple possible 'correct' assignments to the biome, feature, and material fields; the fields are not hierarchical, limiting logical organization; and the primary ecological factors differentiating microbial communites are not captured. In

© Sachs J et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

response to these challenges, the EMP team worked with the ENVO team to devise a new hierarchical structure, the EMP ontology (EMPO), that captures the primary axes along which microbial communities tend to be structured (host-associated or not, saline or not). EMPO is an application ontology, with a formally defined W3C Web Ontology Language (OWL) document mapping to existing ontologies, enabling reuse by the microbial ecology community.

Ecobiomics is a joint project of multiple GoC departments and involves the complete workflow, from sampling in a variety of aquatic, soil, and benthic environments, through sample prep, DNA extraction, library prep, sequencing, and analysis. In contrast to the EMP—where some of the samples and metadata had been collected before the establishment of the MIxS standards—the Ecobiomics project has been able to create metadata profiles for each sub-project to conform to, extend, and build, upon the existing MIxS standards.

Despite these two different contexts, EMP and Ecobiomics encountered a number of common issues that prevented a complete implementation of MIxS. These issues include ambiguous term names and definitions; inconsistencies amongst the environmental packages; non-standard ways of dealing with units; and a number of issues surrounding ENVO (the Environment Ontology), which is required for filling out the mandatory MIxS fields "Environmental material", "Biome", and "Environmental feature". We will describe these issues, and, more generally, the successes and challenges of our implementations.

Keywords

Metagenomics, metadata, MIxS, eDNA, amplicon sequencing

Presenting author

Luke Thompson and Joel Sachs

References

- Buttigieg PL, Pafilis E, Lewis S, Schildhauer M, Walls R, Mungall C (2016) The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. Journal of Biomedical Semantics 7 (1): . https://doi.org/10.1186/s13326-016-0097-6
- Gilbert JA, Jansson JK, Knight R (2014) The Earth Microbiome project: successes and aspirations. BMC Biology 12 (1): . https://doi.org/10.1186/s12915-014-0069-1
- Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, G Chain PS, Charlson E,

Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, Gil IS, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpides N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E, Methé BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Pruesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spor A, Stombaugh J, Tiedje JM, Ward DV, Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T, Glöckner FO (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nature Biotechnology 29 (5): 415-420. https://doi.org/10.1038/nbt.1823